OPERATIONAL DEFINITIONS OPERATIONALLY DEFINED

STUART C. DODD

STUART C. DODD

OPERATIONAL DEFINITIONS OPERATIONALLY DEFINED

STUART C. DODD

ABSTRACT

This paper attempts to meet the challenge of defining operational definitions operationally. A definition is operational to the extent that it specifies the procedure for identifying or generating the definiendum and finds high reliability for the definition. The logical form of this definition, its gradational phrasing, the concepts of "procedure" and "reliability," and the two types of operational definitions are commented upon. The importance of reliability for scientific work is stressed. Experimental procedures for measuring the utility, the reliability, the validity, and the usage of any concepts defined are suggested and proposed as criteria for the excellence of any sociological definition.

Those sociologists who advocate greater use of operational definitions have been challenged to define "operational definition" operationally. This paper attempts to meet that challenge. It amplifies Lundberg's discussion in the March issue of this journal and assumes affirmative answers to the questions he puts (p. 741) as to the points of agreement between operationists and their critics.

The proposed operational definition of operational definitions is:

A definition [*genus*] is an operational definition [*species and definiendum*] to the extent that the definer (a) specifies the procedure [*differentia (a)*] (including materials used) for identifying or generating the definiendum, and (b) finds high reliability [*differentia (b)*] for his definition.

Less rigorously stated, the differentiae of operational definitions include reliably specified procedures.

1. Our first comment is that this definition is in the 'modified Aristotelian' form recommended by the Subcommittee on Definitions of the Committee on Conceptual Integration in that it states, adequately for the definer's purpose, the genus and the differentiae of the species, of which the definiendum is a member. The purpose here is to contribute to the general function of science—namely, the predicting and controlling of phenomena.

2. This definition is itself an operational one to the extent that the two following operations are carried out: (a) The specifying of procedures, including materials used. An operator must state these procedures in adequate detail in order to generate an operational definition. (b) The finding of the reliability index for the definition. This means experimentation and computation of a statistical index of reliability. An operator must determine this reliability in order to identify the definition as operational to the extent that the reliability index that is observed approaches its maximum.

3. The phrasing of this definition is in gradational form rather than in the usual "this" or "not this" all-or-none form of the dictionaries. This form permits the grading of definitions from nonoperational to highly operational without an arbitrary boundary line where they suddenly jump from nonoperational to operational. When any series is divided into only two class intervals, such as "all" and "none," it is less accurately specified than if subdivided into more class intervals. Allowing for gradations of operationality means that almost every definition may be considered as slightly operational in that it involves the operation of specifying the differentiae. Whether this extent of operationality is adequate depends on the reliability index. As this index approaches unity it measures an adequately

specified definition, however slight the specifying of procedures in it.

This gradational phrasing of the definition helps to dispose of the "dichotomizing fallacy" that perennially plagues all human thinking. Among sociologists one currently prominent dichotomizing fallacy is the distinction between the qualitative and quantitative. These can be reconciled as a continuum or graded series with the qualitative as one limit of the quantitative. Every conceivable quality (whether element or configuration, whether physical object or abstract value, or any other kind of knowable entity) can be conceived as either present or absent—i.e., as an all-or-none variable. It is a primitive quantity of zero or unity, the unit being that quality whether war, laughter, a person, truth, or anything else. Every quantity is a quantity of some kind—i.e., of some quality. Every quality can thus be conceived as the lower limit of precision of a quantity. At the lower limit of precision the number of gradations is just two— i.e., the quality either exists or does not exist in the situation studied. A quality becomes more precisely quantified when it is observable in more than two degrees —as, for example, in the four degrees implicitly involved in comparing adjectives and adverbs, the negative, the positive, the comparative, and the superlative degrees ("none," "some," "more," "most"). A quality is most precisely quantified whenever it can be expressed as a multiple of some cardinal unit. Cardinal units are equal and interchangeable, whereas ordinals merely express a rank or sequence without asserting equal intervals between "first," "second," "third," etc. These varying degrees of precision are distinguished by appropriate scripts in dimensional analysis, which is thereupon able to combine them in

the same formulae and equations and solve for either unknown quantities or unknown qualities. The gradational phrase "to the extent that" in the definition above allows for any current degree of precision from the usual and least precise all-or-none degree of quantifying to scales of cardinal units. This gradational phrase thus covers the purely qualitative differentiae and the most precisely quantitative differentiae and every gradation between them.

4. There are two subspecies of operational definitions and these two may overlap. They are the identifying or testing type and the generating or creative type. An operational definition states the procedures which must be carried out in order either to identify a case of the definiendum or to produce it. As examples of operational definitions, a kitchen recipe states the operations for generating a cake, while a Chapin-Leahy scale states the operations for identifying the socio-economic status of a family; a formula for a mean or for any statistical index states in algebraic symbols the operations for generating that index, while instructions for examining a passport state the operations for identifying the holder's nationality; the details of the law and supplementary administrative regulations for licensing motorists state the procedure for generating a "licensed motorist" and, when applied in retrospect, can identify whether a given person is a "licensed motorist" or not. Such recipes, scales, formulae, instructions, rules, all specify the procedures to be used upon specified materials in order to secure or to be sure one has secured that which is defined.

5. Since the concepts of "procedure" and "reliability" are the crux of an operational definition, they may in turn be defined. For the present purpose of clari-

fying operational definitions, a procedure may be defined as any human action (genus) to the extent that such action is a means to ends (differentia *a*) which is communicable by the actor (differentia *b*). The operational differentiae implicit in this phrasing may be explicitly stated as: "Get a person to communicate the actions which he uses as a means to his ends. Such communicated purposeful actions are called 'a procedure.' " The differentia of *communication* is a behavioristic test of whether the action is purposeful, an intended means to an intended end, or is just an accidental sequence into which an outsider reads a purpose. Communicating makes more objectively observable the subjective purposing and lessens the controversy over whether it involves consciousness of the means and of the ends. Moreover, communicating an action tends to make it more definite, formal, and repetitive, and these are connotations of the concept "a procedure." Substituting this definition for the term "procedure" in the operational definition of operational definitions above yields the paraphrase: "A definition is operational to the extent that the definer *a*) communicates the actions to be done as means of identifying or generating the definiendum, and *b*) finds high reliability for it." An operational definition is thus any statement, whether as brief as a sentence or as long as a book, which reliably tells what to do, first, second, third, and with what ingredients, in order to test for the presence of, or to produce, that which is defined.

6. "Reliability" may be briefly defined as any index (genus) measuring the degree of agreement (differentia *a*) among reobservations of the same phenomenon (differentia *b*). Unreliability is the lack of such agreement, or variation among reobservations. In more semiotic language, a sign-vehicle, such as a concept, is reliable in proportion as the designata are constant for all interpretants under specified conditions. A reliable concept is one whose referents are standardized for all users of the concept. The degree of reliability is measurable by some appropriate statistical index. Thus the exact operational definition of reliability is stated by the formulae under this heading in a statistical textbook. There are formulae for measuring the degree of agreement within the sample observed among reobservations of a constant (i.e., a single-valued quantity), such as a difference of means $(M_2 - M_1)$ and among reobservations of a variable (i.e., a many-valued quantity) such as a reliability correlation $(r_{1I} = \Sigma z_1 z_I / N)$. There are further reliability formulae for estimating the probability of any assigned degree of agreement among reobservations within the universe sampled. These formulae involve standard-error formulae if for large samples, and fiducial-limit formulae if for small samples. Thus the definiendum may be the "socioeconomic status of farm families" as defined by the Sewell scale. A second application of this scale to a sample of families then yields a reliability correlation coefficient measuring its variable errors and a difference between means of the two applications measuring its constant error. These are reliabilities within the sample and when divided by their standard errors yield estimates of probability which measure reliability in the universe sampled. This procedure is familiar to most scientifically oriented social scientists today. But what is new is that definitions of concepts which denote a class—i.e., a qualitative kind of entity—with no apparent quantification can have their re-

liability similarly determined by a statistical index derived from controlled experimentation.

Reliability formulae are not limited to determining the reliability of quantities, since appropriate formulae exist for determining the reliability of qualities as well. One procedure is to collect and record many items which are referents of the concept-to-be-tested and many other items which are similar in varying ways but are not referents of that concept, in the judgment of the collector. The accuracy of the collector's judgment is not important so long as the collection includes instances of that concept and other borderline instances. Let competent persons classify these items independently as included or excluded under the concept according to the definition of it to be tested. Compute the percentage of agreement, or identical classification of the items, by these independent persons. This percentage is one index of the reliability of the concept as defined and can be compared with the degree of reliability of any rival definition when applied to the same items by the same persons. A formula for this reliability is $\%_A = 100 n_i/n \ (\pm \sigma\%)$, the percentage of agreement ($\%_A$) is the number (n_i) of identically classified referents divided by the total number (n) of referents in the sample collected, multiplied by one hundred; plus or minus its standard error ($\sigma\%$) if it is desired to generalize from this sample.[1]

The last paragraph containing an operational definition of the reliability of a qualitative concept seems so new an application of reliability principles as not yet to be found, to the author's knowledge, in any textbook of statistics or of social research. It belongs under experimental semantics—if there be such as yet. Although almost unthought of among sociologists at present, it should become one of the most basic and often used technics for sifting the concepts used in any serious research. An example of its use occurs in the measurement of the reliability of the operationally defined system of concepts in the author's *Dimensions of Society*.[2] Here some seven hundred odd concepts are defined in algebraic equations which specify the mathematical procedures to be performed on the symbolized entities to obtain the concept in question. Whenever the symbolized entities have specified the procedures by which they were secured from phenomena, these definitions become more fully operational—but still only in proportion to their reliability. All these formulae were compounded from some sixteen basic concepts, the reliability of which was experimentally measured as follows: Five hundred sets of data from all the social sciences were representatively sampled to serve as cases of referents for any systematic sociology. Two persons independently applied these basic concepts to this body of referents in writing for each referent set of data a

---

[1] Note that this observed reliability is relative to the sample of referent items used. A different selection of items might change the reliability index. Thus a larger proportion of borderline items would probably lower the reliability observed. This means that the sample of referent items should be standardized in a publicly available record of those items so that any other investigator could use that identical sample and thus keep constant conditions for the reliability experiments. This dependence of the index upon the sample selected is comparable to the dependence of a correlation upon the range of its population. Thus a correlation of 0.5 in a one-year age range can be increased to a correlation of 0.76 in another age range where the sigma is doubled. Since correlation coefficients are indices expressed in standard deviation units, they are comparable only to the extent that their ranges are comparable.

[2] New York: Macmillan, 1942.

descriptive formula compounded of the basic concepts. The percentage of agreement, or identically written formulae, for these five hundred referent cases was calculated. This reliability percentage ran from 93 per cent to 100 per cent in a series of such experiments under varying conditions. This pioneering study in dimensional sociology demonstrates how reliability indices can be experimentally determined for qualitative as well as for quantitative concepts.

The Committee for Conceptual Integration was organized because our sociological concepts have such shifting designata; their meaning often varies from user to user; they often fall short of the scientific ideal of communicating a standard body of referents. Yet in spite of this realization among the members of the Committee on Conceptual Integration and others, it is amazing to find so much indifference or ignorance among them of the primary importance of determining reliability in any kind of defining of concepts. Of what use for science is an unreliable concept, whatever its excellence in other respects? Scientists in the older sciences know better than to fit theories to observations until those observations are proven to be facts—i.e., until their reliability has been established by reobservation by independent observers. In so far as sociologists use observations, or summaries of them in concepts, which are unreliable further work based on them is largely a waste of time. Improvement of the reliability of our verbal instruments and other symbols is a much-needed emphasis in research today. A prediction may be ventured that the sociological publications with the greater reliability of concepts will tend to have greater longevity. The unreliable concepts will prove more ephemeral.

7. The two differentiae of operational definitions—the specifying of procedures and the finding of high reliability—may vary independently. Any definition may specify procedures, but the specifying may be so subjective or unclear as to result in low reliability when that definition is reapplied to the same sample of referents by another person. Conversely, a definition may have little or no specifying of procedure (and so not be an operational definition) and yet have high reliability. This result is more frequently possible with simple perceptual terms, as in defining a "trident" as a fork (genus) with three prongs (differentia a). But the more complex and abstract the concept is, the less likely it is to have high reliability without specifying procedures. Critics of operational definitions are hereby challenged to produce a definition of some sociological concept, such as a "social force," which lacks specification of procedure for identifying or generating the force, and yet has a reliability of more than 0.90. The author's operational definition of an effective social force is "all [genus] that accelerates [differentia (a)] a change in people [differentia (b)], measured by the procedure of multiplying the number of people changed (P) by the amount of their acceleration (A); the formula is simply $F = PA$."[3] This definition has been experimentally shown by one of the author's graduate students to

[3] Acceleration is the time rate of change of a process, which in turn is an observable change in time. So acceleration is measured and defined by the amount of change in some index (I) twice divided by time. Therefore $A = I/T^2$ is its dimensional formula.

In case the quantity of acceleration and of people is not determinable but it can be qualitatively asserted that "people have changed" the formula becomes the logical product, $P \cap A$, which is "that which is jointly characterized by 'people' and some 'acceleration of change.'" In dimensional sociology, the zero exponent denotes a quality, a logical class, and the formula for this qualitatively asserted or unquantified force is $F^0 = P^0 A^0$.

have a reliability greater than 0.95, which is near the maximum or perfect reliability of 1.00.

For a definition to be an operational one, each of the two differentiae are necessary conditions and together they are sufficient conditions. This statement may be questioned by someone who believes that "specifying procedures" is enough to differentiate operational definitions from other definitions and that "finding high reliability" differentiates another species of definitions—namely, "reliable definitions." Our contention, however, is that "an operational definition" has come to mean both. "Specifying procedures" would have no superior intrinsic merit worth quarreling over compared with other kinds of differentia in a definition, were it not that by specifying procedures (including always the materials involved in those procedures) greater reliability of definitions is achieved. Physical scientists have overwhelmingly found this to be true. Social scientists are increasingly discovering its truth. This last statement need not remain the opinion of an operationist merely; it can be experimentally verified by measuring the reliability indices of a set of concepts when operationally defined as compared with their nonoperational definition. Here is an opportunity for some graduate student to make crucial experiments in sociological methodology. Operationalism would have few advocates did not those advocates see in it a technique for making concepts more reliable, for standardizing their referents, and thus a technique for getting out of the conceptual morass which occasioned the formation of the Committee on Conceptual Integration.

It is unfortunate that the growing interest in operational defining has centered on the "specifying of procedures," due to the label "operations," and has neglected the far more important aspect of testing reliability. The operations are but a means to the scientific end of prediction and control. To the naïve operationist the greater reliability is assumed, or considered a connotation of "the operational," and is therefore not adequately communicated to the critic of operationism, who naturally then sees no magic in merely "specifying procedures." If high reliability is explicitly denoted by the term "operational," as in the definition proposed here, this important property of definitions will be better communicated and develop more consensus in the controversy over operational defining.

In this connection should be mentioned the fallacy of assuming that operationists are concerned *merely* with the clearness and preciseness of terms, and are less concerned with the "organizing ability and utility" and "meaningfulness" of the concepts.[4] This is a preposterous assumption in view of the fact that the concepts to be defined are set by the theory we adopt, and are therefore usually the same for operationists as for others. Operationism is not itself a sociological theory. It is merely a method of attacking a problem faced by all scientists, namely, defining the concepts employed, *whatever* these concepts may be. The utility of these concepts is another question, in which operationists are as much interested as anyone.[5] What is more, operationists have faced the crucial fact that the only way to determine the relative utility of different concepts is first to define them with reliability.

[4] See Harry Alpert, "Operational Definitions," *American Journal of Sociology*, XVLII (May, 1942), 981.

[5] See, e.g., G. A. Lundberg, *Foundations of Sociology* (Macmillan, 1939), chaps. v–vii.

8. The *reliability* of a concept is to be clearly distinguished from its *validity*, its *utility*, and its *usage*. Psychologists dealing with tests say that "reliability" means "how well the test measures whatever it measures," while "validity" means "how well it measures what it claims to measure." Operationally, the reliability correlation coefficient between two administrations of a test defines its degree of reliability, while the validity correlation coefficient between the test and an accepted criterion of whatever the test claims to measure defines its degree of validity. Thus the validity of an intelligence test is its degree of correlation with some currently and widely accepted indicator of intelligence, such as school marks, occupational achievement when opportunity has been equal, presence in a home for feeble-minded vs. graduating from college with honors, etc. The criterion is usually less precise and more costly or time consuming than the test so that, if the validity is high enough to justify substituting the test, a gain in precision and economy has been achieved. Also, since the test, once it is validated, may be given to people before the criterial behavior takes place, it promotes the prediction and control of that behavior.

Validity always involves a criterion. Without an accepted criterion, validity in the technical sense accepted in psychology and statistics and described here has no meaning. Furthermore, validity, when determined, is relative to that specific criterion and may have a different value with respect to another criterion. The validity correlation is the proof of the extent to which a new and more efficient indicator of some phenomena can be substituted for a less efficient but conventional and familiar indicator of those phenomena.

The *utility* of a concept for scientific purposes means how well it contributes to our ability to predict and control phenomena. In the long run scientists find a concept useful or useless in proportion as it enables them to understand and hence to predict and control relevant phenomena better than with alternative concepts or absence of them. Thus "oxygen" with its denotata proves more useful than "phlogiston"; "behavior" supplants "consciousness" as the more useful term in psychology; and "correlation" enabling prediction via its regression equation displaces the vaguer concept of "concomitance" popular in John Stuart Mill's day.

Utility requires time for a consensus to develop among scientists. It is not often measurable, in current sociology at least, as neatly as reliability or validity. Conceivably experiments could be set up, however, which would measure the relative predictive efficiency of alternative concepts, or alternative definitions of one concept. The instructions for such an experiment would constitute the operational definition of the "utility of a concept." At present, with inadequate specifying of procedure, the "utility of a concept" may be defined as the correlation coefficient between the use of that concept and the efficiency of predicting its relevant phenomena. An operator would thus have to collect instances of the concept's use and nonuse, together with some estimate of the resulting degree of efficiency in predicting in each instance. ("Prediction" is here used as including "control," "control" being that subclass of "prediction" where man's actions are a factor in bringing about the predicted outcome.)

The *usage* of a concept refers either to the number of people using it, or using it with specified referents, or with a specified definition. Thus some sociologists

use the concept of "culture" as including animal phenomena that are similar to human culture, while others use it as excluding such animal phenomena. The operational definition of "usage" would be to measure the proportion of a specified plural which uses the concept or uses it in a specified way (whichever was at issue).

It is possible to take usage as a criterion for validating a concept. One procedure for this validation would be for a representative and authoritative panel of specialists on "culture" to classify each recorded case in a sample collection of several thousand cases as "cultural" or "noncultural" on the undefined basis of their customary use of the concept "cultural." These same cases would then be classified again in all-or-none fashion as "cultural" or "noncultural" on the basis of a specified definition of culture by a number of competent persons (whose classifyings would be averaged for greater reliability). The four-point correlation coefficient would then be calculated between the all-or-none two-point variable of the panel's usage and the two-point variable yielded by the definition. This correlation defines the validity of this definition by the criterion of this panel's usage. It could be compared with the validity of any other definition of the same concept by comparing correlations with this criterion. Of course, this illustration assumes that current usage is a worth-while criterion by which to validate a concept, whereas this assumption may not be at all defensible.

Of the four properties of concepts defined above, *utility* would seem the most important for science, with *reliability* next, while *validity* and *usage* may be currently desirable but would seem less important for scientific progress in the long run. The excellence of any definition of a concept, in addition to its logical form, might well be gauged by these four properties. Whether a definition is operational or not seems to us a partial test of its excellence, but more rigorous tests are its correlations showing how useful for prediction, how reliable, how valid by specified criteria, and how widely used the concept is.

AMERICAN UNIVERSITY OF BEIRUT